

## FAO Standard Seed Security Assessment

## BASIC STEPS IN RUNNING ANALYSIS WITH PIVOT TABLE in MS Excel 2003 ou XP et MS Excel 2007-2010

## Day 4: Session 9

**Step 1:** Open MS Excel program and open a file or database that is already managed and cleaned.

**Step 2:** Go to the Menu and look for “Insert” tab. In the Insert tab you will find a Table group. Click on PivotTable and then on PivotTable as shown in **Error! Reference source not found.:**

## Window 1

Database - Rcn follow up survy1 - [Compatibility Mode] - Microsoft Excel

File Edit Format Tools Data Review View Window Help

Insert PivotTable

Sort PivotTable

Summarize data using a PivotTable

Ques ID

1.036

22.085

20.031

17.085

17.019

1.038

17.015

6.033

20.036

21.078

1.015

1.022

17.018

20.034

20.035

20.040

21.067

1.020

1.025

6.040

22.048

22.042

5.001

20.041

17.055

14.119

17.030

1.021

Database - Rcn follow up survy1

RESNAME

GENDER

PROVINCE

DISTRICT

COMMUNE

VILLAGE

MALOVER18

FEMOVER18

MALLESS18

FEMLESS18

CHILDREN

INSOK

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398</

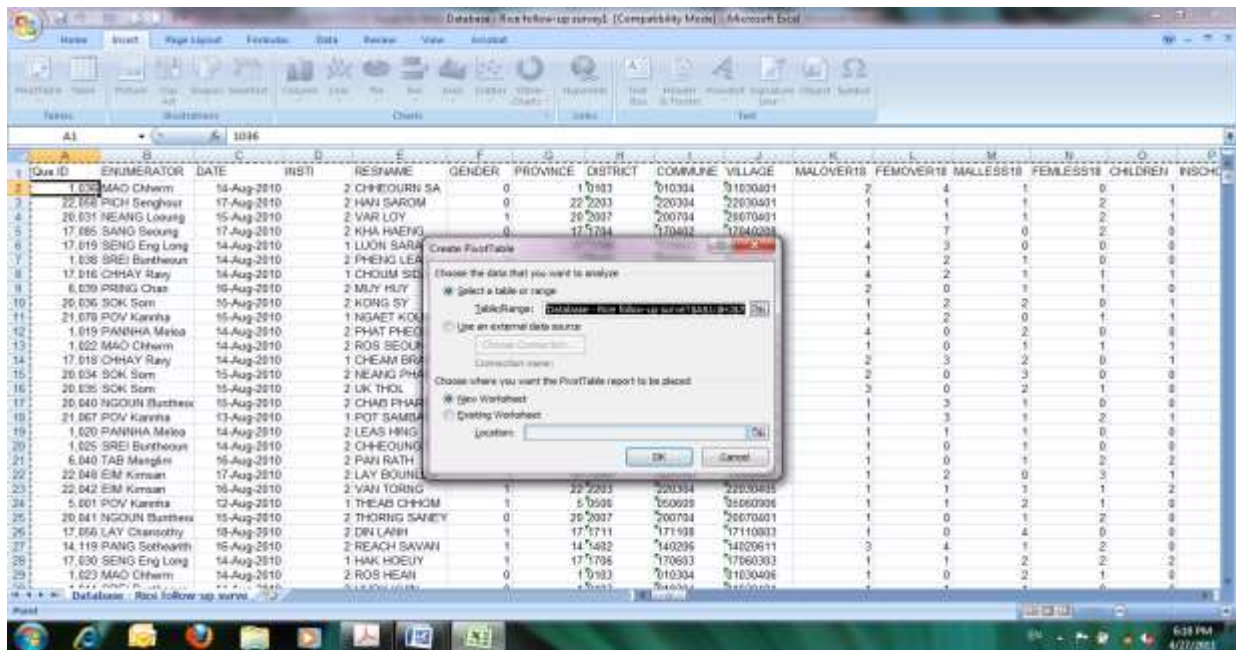
**Step 3:** After clicking on it, Excel will automatically display a dialog box called “Create PivotTable”, please see **Error! Reference source not found.**. Now the program is asking about which data or range of data is to be analyzed. Automatically, if your database is completely cleaned and ready for analysis, the Excel will select all of the data available in the worksheet. .

The displayed dialog box provides the following options:

**“Choose the data that you want to analyze”:** the available options are as following

- ⦿ “Select a table or range”: this is asking what data available in the worksheet to be selected for analysis.
- ⦿ “Use an external data source”: this option enables us to use data from outside e.g. Another server.

## Window 2



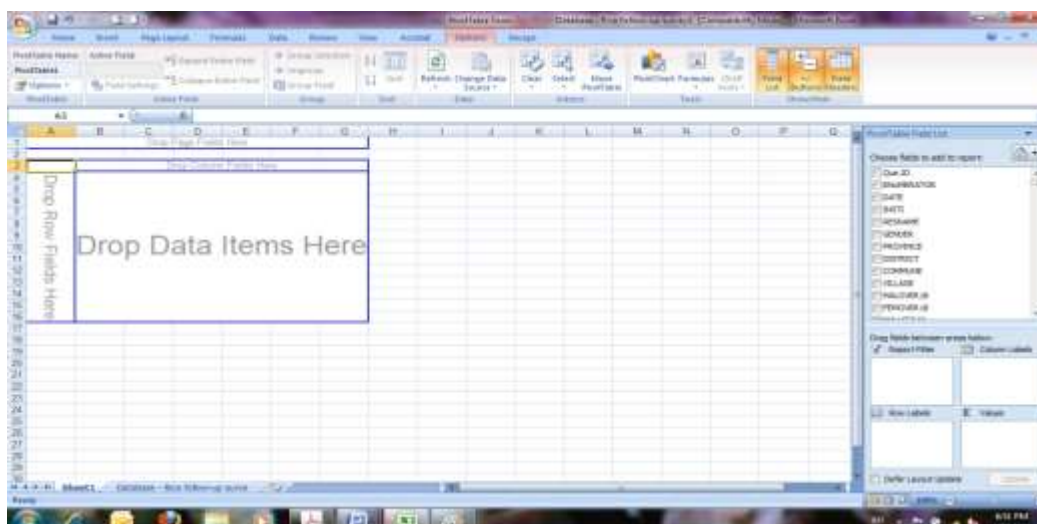
**Choose where you want the PivotTable report to be placed:** This is to ask us where to put the produced PivotTable. The available options are as follows:

- ☉ New Worksheet: if this is selected the produced PivotTable will be placed in a new worksheet, which is newly created automatically.
- ☉ Existing Worksheet: if this is selected the produced PivotTable will be placed in any available / created worksheet. If this option is selected we need to define which worksheet to be used by specifying in the Location box. Row and Column also should be clearly specified.

After completing all of these steps, click on OK button.

**Step 4:** Next, the following Window will be shown, please see **Error! Reference source not found.**

## Window 3



The displayed window is divided into two main parts. In the far right part, we will see a dialog box with the name of PivotTable Field List. This dialog enlists all of the variables that we have selected from Step 3 and these variables will be used for the any analysis.

- Note: Variables available in the PivotTable Field List are the name put in the very first row in the worksheet.

Below the same dialog box, we will see a phrase named “Drag field between areas below”. This is where we can take any variables from the above list and put them into analysis and/or calculation based on their types e.g. calculating average, sum, counting...etc

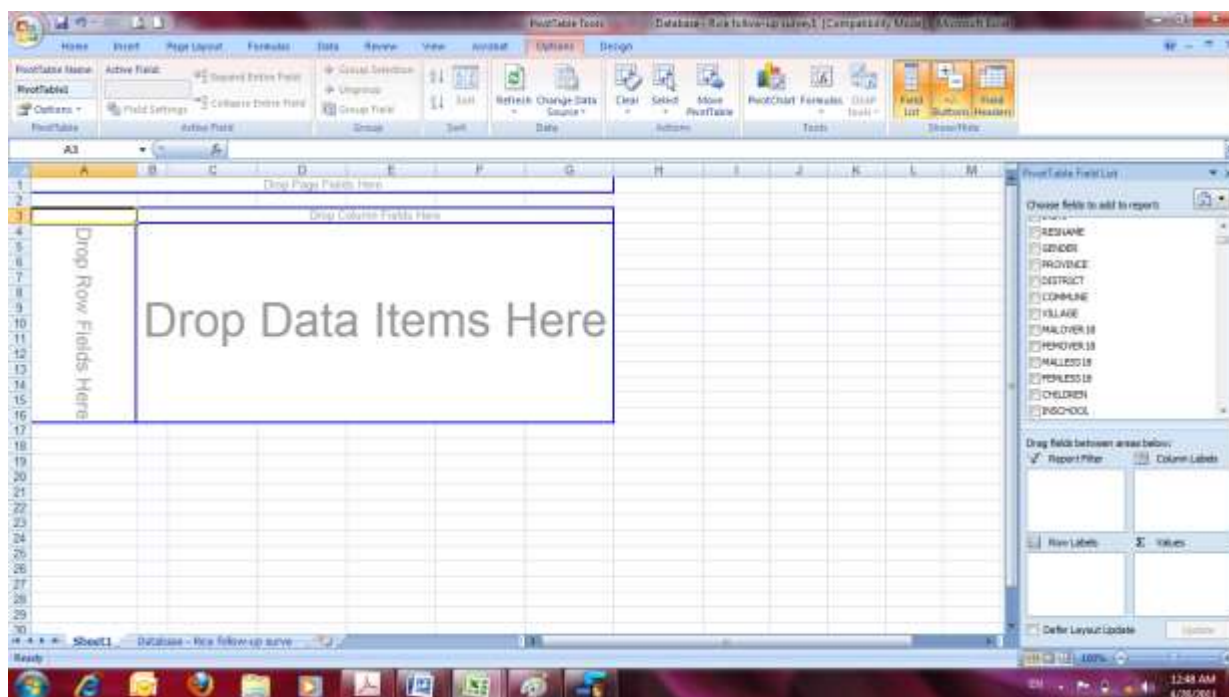
There are two main options to be noticed:

- ∑ Values (at the right bottom of the worksheet) : In case we want to sum up the value from all cases of one specific variable we just have to drag that variable and put it into this box. Automatically, Excel will do the calculation of summing for that variable.

OR

- Drop Data Items Herein (at the left upper side of the worksheet), we just have to select the variable whose values are supposed to be analyzed and drag it onto the area named “Drop Data Items Here”. Please see **Error! Reference source not found.**

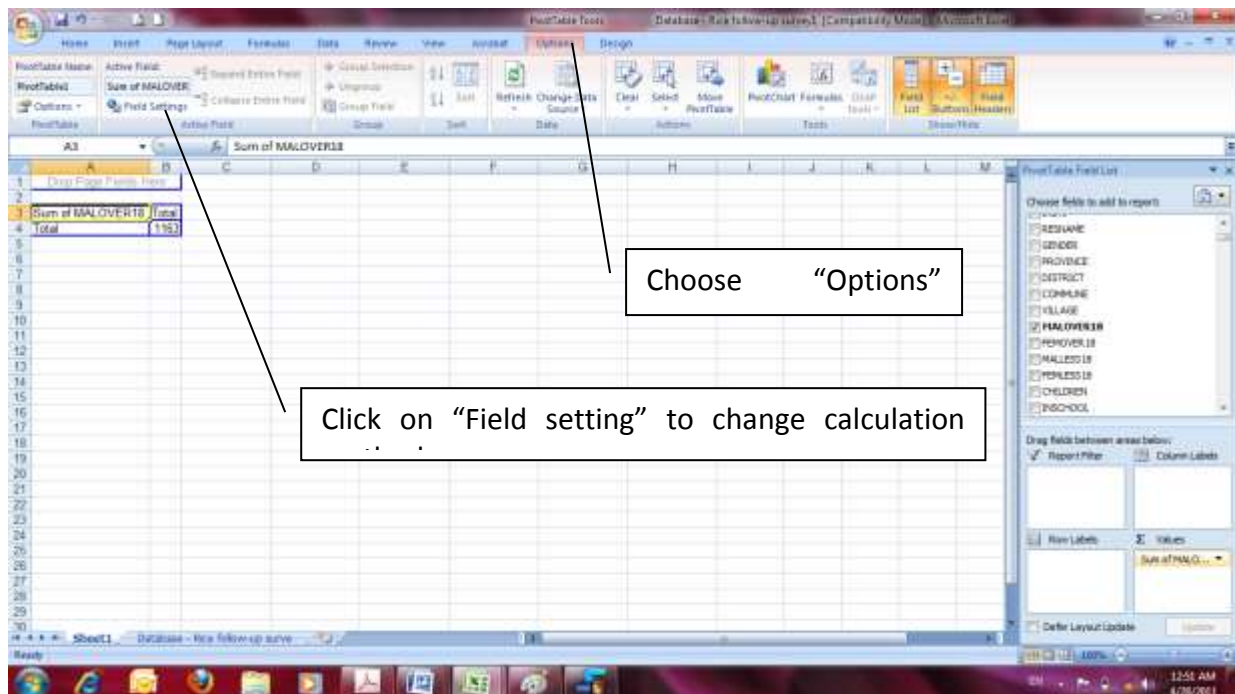
#### Window 4



As default the sum of the specific variable is made; in case we want to change the analysis from a sum to another calculation, as count number, max, min, StdDev, etc.), follow the instruction below

- a. After completing the dragging of variables onto either “ $\Sigma$  Values” or “Drag Data Items Here” please go to Menu and find “Option”. In the group of “Active Field”, click on “Field Setting”. See **Error! Reference source not found.**

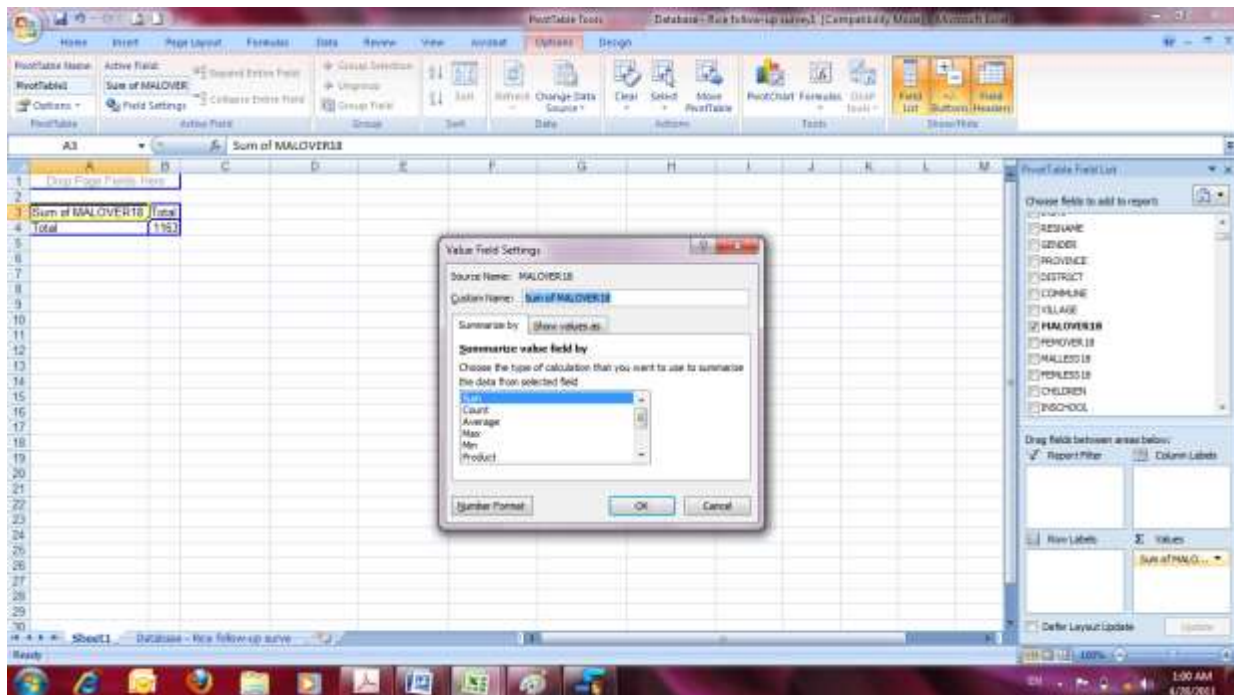
Window 5



- b. The next dialog box will appear with the name as “Value Field Settings”, please see Window. Here is to choose any calculation method you would like to run and then click on “OK”. The calculation will be done automatically.
- c. Alternatively, click on Variable name in the “ $\Sigma$  Values” field setting and the “Value field setting” the dialogue box will appear. Here is to choose any calculation method you would like to run and then click on “OK”. The calculation will be done automatically.



## Window 6



**Row Labels:** In case we want to calculate one variable by classifying the result based on another variable, we will have to drag the variable that is used for classifying onto the “Row Labels”. For instance, we want to find out the average of land holdings of farmers classified by their sex (between men and women). To do so, we need to drag the SEX variable into the “Row Labels” and then drag the variable on landholding into  $\Sigma$  Values or drop it into “Drag Data Items Here”. In general, the variables used for grouping or classifying are the types of Nominal or Ordinal variables.

**Column Labels:** this is similar to Row Labels. The difference is that the result from “Row Labels” is to classify and appear in row, while from “Column Labels” is to appear in column.

**Report Filters:** this is also similar to both “Row Labels” and “Column Labels”. The difference is that the result from this “Report Filters” is to appear by allowing us to select on only any specific options for display.

### 4.7.2 Description and use of the common data analysis procedures

As in the questionnaire development, it is important to remember how each type of response can be used during the analysis. The main options are:

*Yes / No questions coded as y/n or as 1 and 0:* There are two ways to go during the analysis, depending on whether the code is y / n or 1 and 0. If coded 1 for yes and 0 for no, first calculate the sum and calculate later the percentage of 0 and 1 out of the total sample size. For the y and n types of entry, first drag variable heading to the column label” and then to  $\Sigma$  Values and then calculate the counts (frequency) for each.

a) *Single response out of a given list:* These results are mainly used as differentiating parameters which will be placed either in the row or column section of the table layout. In order to get

their frequency (count numbers), drag the questionnaire number to  $\Sigma$  Values, then click on the field setting and change to count numbers. table.

- b) Quantitative data: These results are mainly used as data to conduct some calculations in the data section of the table layout. The most common used analysis types are: averages, maximum, minimum, standard deviation, etc. They are also the main datasets which are used deriving additional variable such as yield, total household members, grouping of household by land size. Grouping of households according to land size can be done directly in the database as an additional parameter. Once the classification has been give the codes in the right variable column, then the analysis can be run as follows;
  - i. Adjust the data range and update the dataset first
  - ii. The variable (land size category) to be analyzed dragged into the corresponding Row or column field setting to set the pivot table for analysis.
  - iii. The variables to be analyzed is again dragged into  $\Sigma$  Values field to run the analysis required, in this case questionnaire number and field settings Count (for frequency) and maintain the other specific parameters as they were before as they will be calculated to the specific differentiating variable accordingly.
  - iv. A differentiating variable such as location, agro-ecological zone, or household type could be placed in the corresponding column field settings in order to analyze the data by location or agro-ecology or household type or a combination of two.
- c) Qualitative data 1– coded single response with Others option (specify)....: Here predictable responses are normally coded in the questionnaires beforehand, while giving allowances for unpredictable responses as others (specify)..... Here, others have to be coded before the analysis is done. This can be done by the data clerk in consultation with the database manager. The obtained new codes are listed and continued numbered and thereafter the analysis can be done. They can be either used as a differentiating parameters (in column or row) while at the same time the variable questnumber is dragged into  $\Sigma$  Values field to generate count numbers (frequency).
- d) Qualitative data 2 - from open ended questions. Open ended questions tend to produce a range of qualitative data. Here also the data clerks in consultation with the data manager have the code all the possible responses before the analysis starts.
  - If the open ended question is a single response, then only one variable entry is possible per questionnaire and a single column is provided.
  - If the open ended question is a multiple response type, then all the possible responses becomes variables and needs different columns.

These are also analysis as in qualitative 1. The obtained groups (code list) are used as differentiating parameters (in column or row) while at the same time the variable questnumber is dragged into  $\Sigma$  Values field to generate count numbers (frequency).

#### 4.7.3 Use of differentiating (classifying) parameters

The differentiating parameters such as locations, agro-ecologies, household types or seed sources are dragged into the sections of columns or rows in the table layout (1-2 at a time, 3 possible but will generate complex tables). In the example in table 4.3 below, the type of households is the differentiating parameter, and the variables such as number of men or the number of animal units are the analyzed variables.

Running analysis:

- i. Here the differentiating parameter (household type) is normally dragged into a Column or Row field settings,
- ii. The variable (number of men) to be analyzed dragged into the corresponding Row or column field setting to set the pivot table for analysis.
- iii. The variable to be analyzed is again dragged into  $\Sigma$  Values field to run the analysis required, in this case average number of men.
- iv. Ensure that the Value field setting is set to the right type of analysis in this case average.
- v. Where an average is calculated, a standard deviation analysis can be done to see if the averages are significantly different from overall the mean.

Table 4.3 Use of differentiating parameters

Parameter	Household types			Overall result	
	Residents	Returnees	IDP	Average	Std Dev.
Average of men	1.4	1.39	1.45	1.42	0.15
Average of animal units	7.22	5.67	2.6	4.7	0.86

From the above example (Table 4.3) the average number of men per family is not significantly different among the various categories (residents, returnees, IDP). The type of household (differentiating parameter) therefore has no influence on this average and the overall average is the one to be reported. However, there were significant differences recorded in between the average number of animals among the various categories. Therefore in this case, the type of household is a significant differentiating parameter for this parameter. This process has to be done for each of the obtained tables.

## DATA ANALYSIS WITH MS EXCEL XP

In the instructions that follow, the windows always relate to the instructions presented in the preceding bullet points.

First open the Excel file that contains the database see Window 7. The database has to have one row with the headings of each of the different columns. Each heading has to have a different name. Make the names as short as possible but still in a way that you can identify easily the information or kind of response entered under this column, which will help to conduct the data analysis process. Each of the other rows below represents the data of a specific questionnaire, where the initial number makes the links between them.

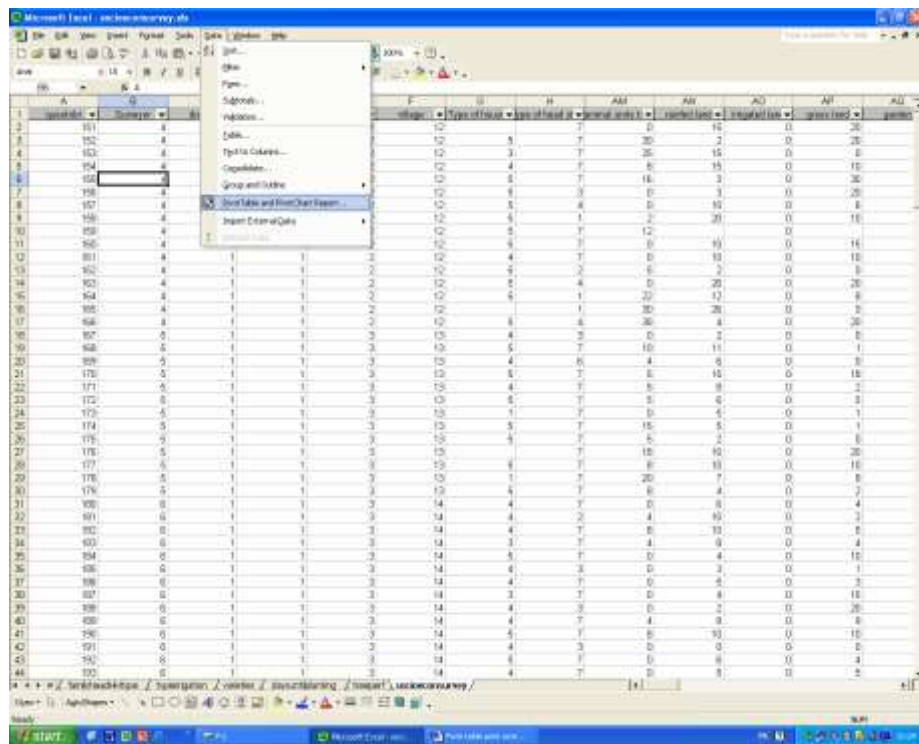
Window 7

questionnaire	sex	district	teleph	union code	village	type of house	type of land	animal units	number of land	irrigated land	gross land	garden
1	151	4	1	1	2	12	7	0	15	0	20	
2	152	4	1	1	2	12	5	7	30	2	20	
3	153	4	1	1	2	12	3	7	25	15	0	0
4	154	4	1	1	2	12	4	7	8	15	0	10
5	155	4	1	1	2	12	5	7	16	3	0	30
6	156	4	1	1	2	12	5	3	0	3	0	20
7	157	4	1	1	2	12	5	4	0	10	0	6
8	158	4	1	1	2	12	5	1	2	20	0	10
9	159	4	1	1	2	12	5	7	12	0	0	
10	160	4	1	1	2	12	5	7	0	10	0	15
11	161	4	1	1	2	12	4	7	0	10	0	10
12	162	4	1	1	2	12	5	2	6	2	0	0
13	163	4	1	1	2	12	5	4	0	20	0	20
14	164	4	1	1	2	12	5	1	22	12	0	8
15	165	4	1	1	2	12	1	30	26	0	0	0
16	166	4	1	1	2	12	5	4	38	4	0	20
17	167	5	1	1	3	13	4	3	0	2	0	0
18	168	5	1	1	3	13	5	7	10	11	0	1
19	169	5	1	1	3	13	4	6	4	6	0	0
20	170	5	1	1	3	13	5	7	5	15	0	18
21	171	5	1	1	3	13	4	7	5	8	0	2
22	172	5	1	1	3	13	5	7	5	6	0	5
23	173	5	1	1	3	13	1	7	0	5	0	1
24	174	5	1	1	3	13	5	7	15	5	0	1
25	175	5	1	1	3	13	5	7	5	2	0	0
26	176	5	1	1	3	13	5	7	18	10	0	20
27	177	5	1	1	3	13	5	7	8	10	0	10
28	178	5	1	1	3	13	1	7	20	7	0	8
29	179	5	1	1	3	13	5	7	8	4	0	2
30	180	6	1	1	3	14	4	7	0	5	0	4
31	181	6	1	1	3	14	4	2	4	10	0	2
32	182	6	1	1	3	14	4	7	8	10	0	5
33	183	6	1	1	3	14	3	7	4	6	0	4
34	184	6	1	1	3	14	5	7	0	4	0	10
35	185	6	1	1	3	14	4	3	0	3	0	1
36	186	6	1	1	3	14	4	7	0	5	0	3
37	187	6	1	1	3	14	3	7	0	4	0	15
38	188	6	1	1	3	14	4	3	0	2	0	20
39	189	6	1	1	3	14	4	7	4	6	0	9
40	190	6	1	1	3	14	5	7	8	10	0	10
41	191	6	1	1	3	14	4	3	0	0	0	0
42	192	6	1	1	3	14	5	7	0	5	0	4
43	193	6	1	1	3	14	4	7	0	5	0	5
44	194	6	1	1	3	14	4	7	0	5	0	5

The Pivot Table feature is an integral part of the MS Excel program, and can be found in the 'data' menu. Select the option Pivot Table Report as shown in Window 8.



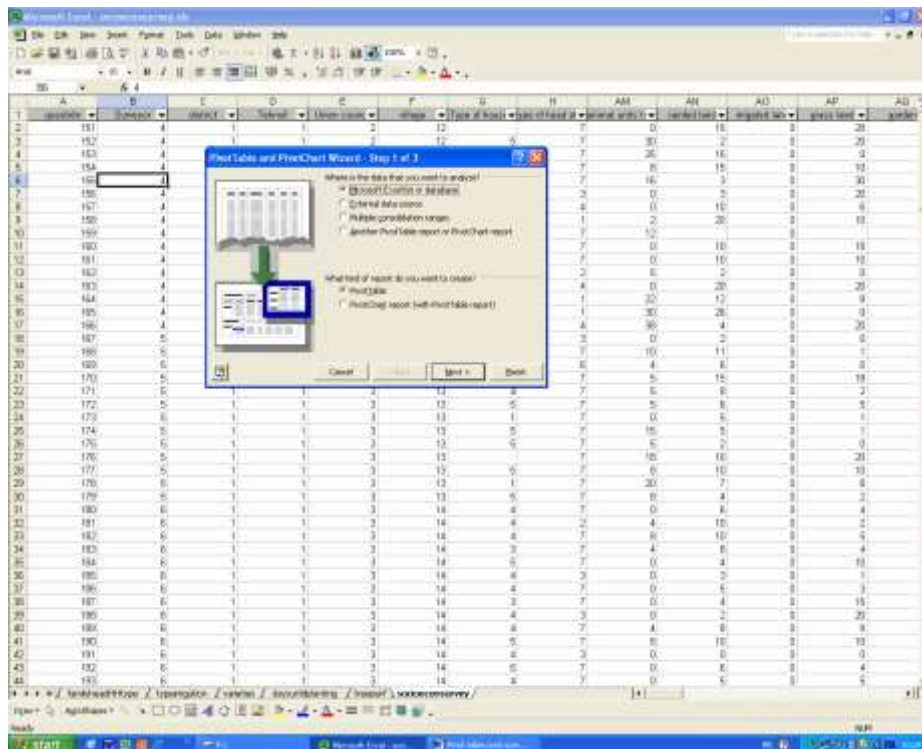
## Window 8



Now follow closely the instructions given below.

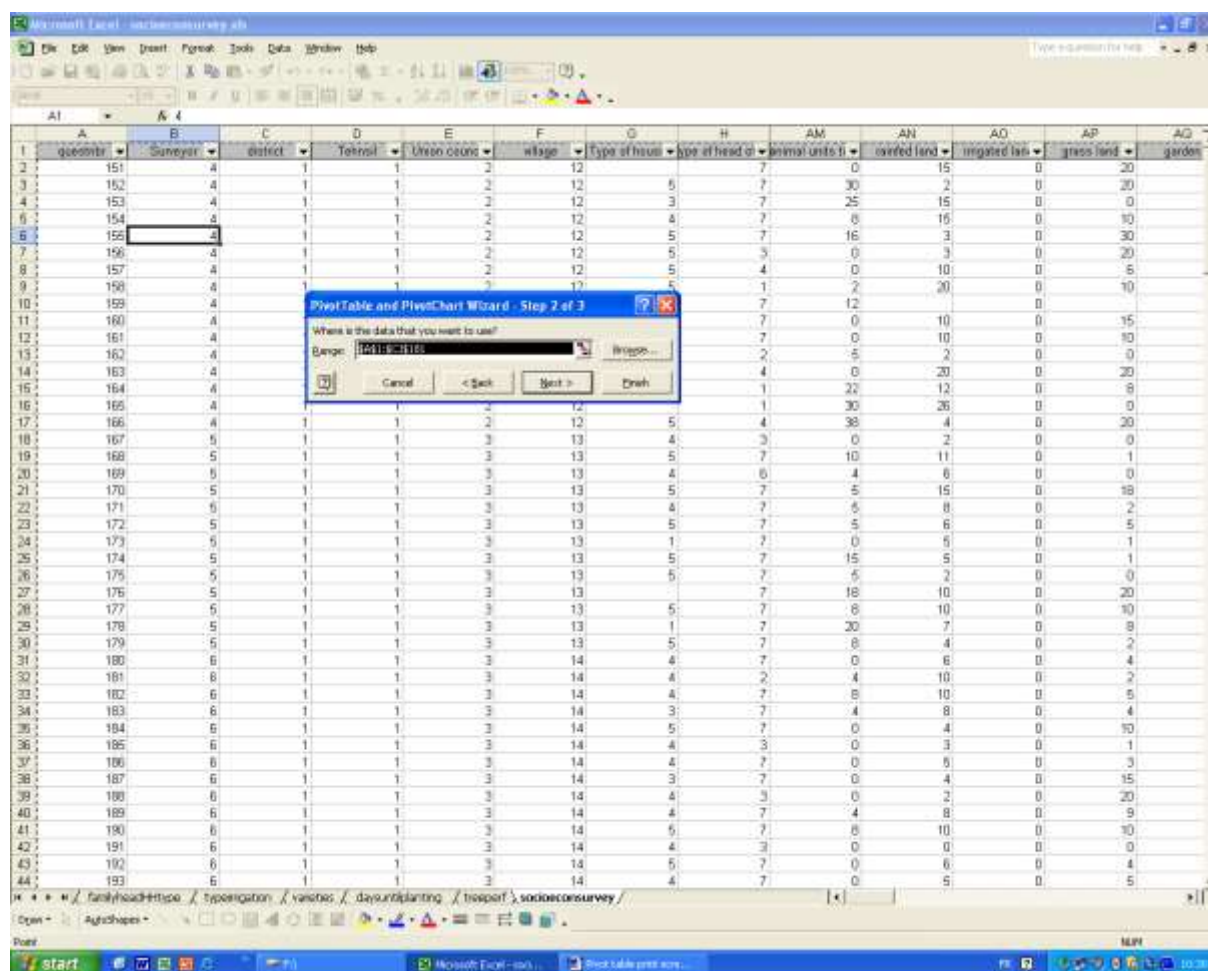
- In the new window, as shown in Window 9, select the option 'Microsoft Excel list or database'.
- Once the selection is made, click 'next'.

## Window 9



- Automatically a new window appears, as shown in Window 10, requesting the user to enter a range where the dataset is placed and should be analysed. The most recent version of Excel selects the range automatically, taking the whole dataset in the previously open worksheet. In previous versions it is necessary to enter the range of the whole table using the cursor and then select the whole range directly in the datasheet.

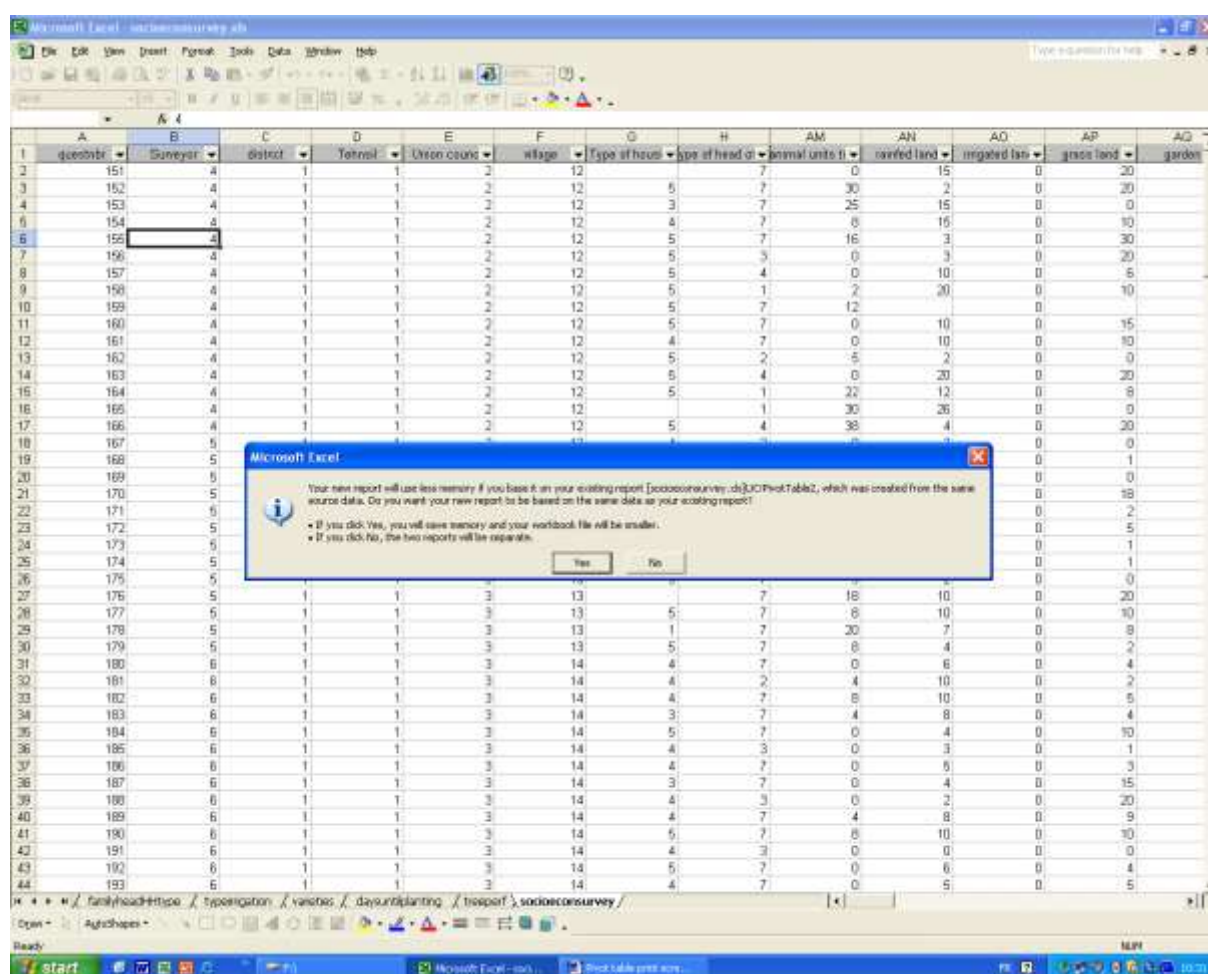
Window 10



After having finished this operation press 'next'.

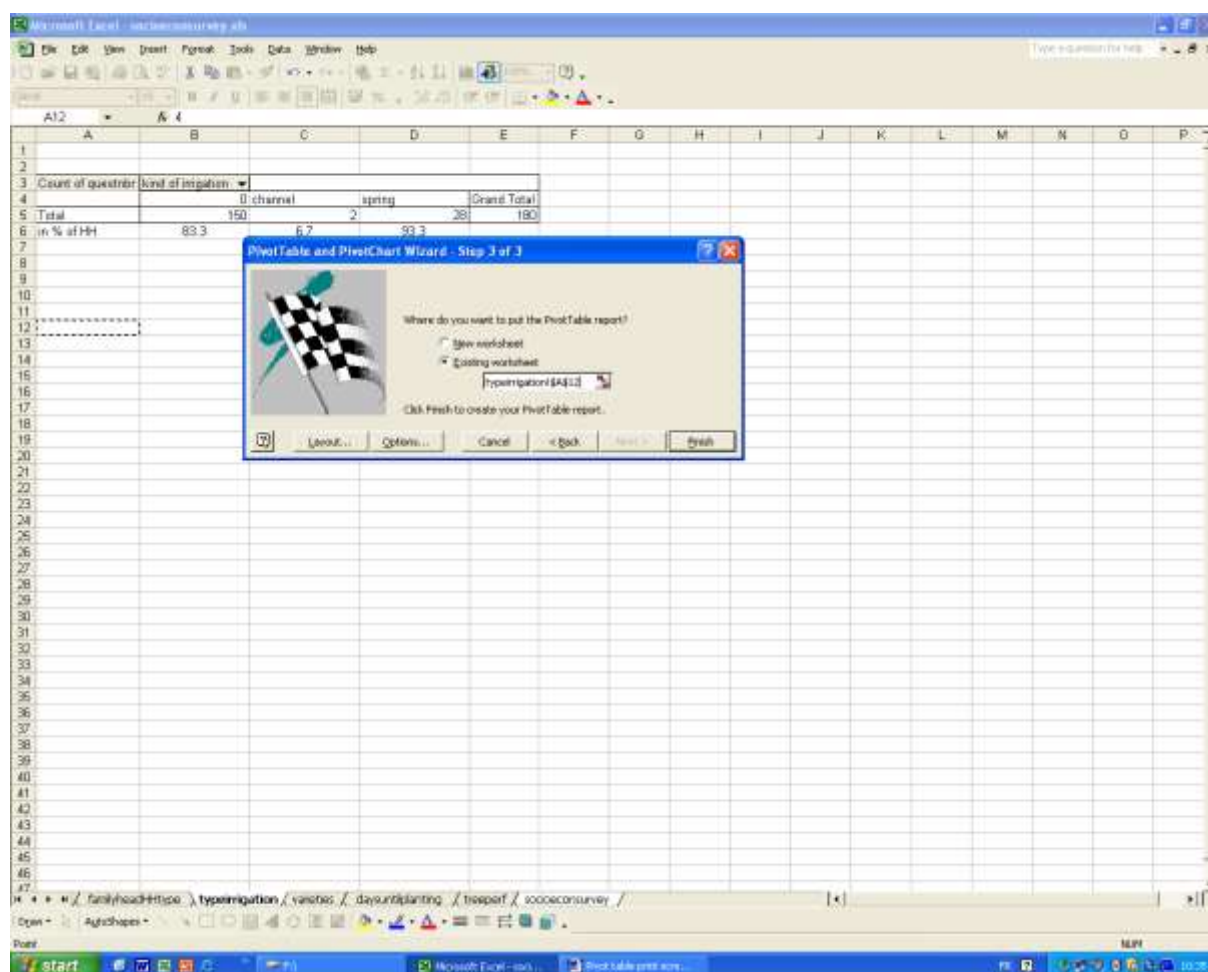
- After the previous step, the next window, as shown in Window 11, asks if the results should be made out of the existing results table or if they should be created using once again the initial data set. This has an implication on the size of the file, which is not a problem, and allows tables to be independent from each other. It is recommended to use the initial dataset, and therefore the 'No' option should be selected. This step is skipped when making calculations for the first time and no results table exists yet.

## Window 11



- In the next window, as shown in Window 12 it is necessary to decide where the results should be stored; either in a new worksheet or in an existing one. If you select an existing worksheet, make sure to enter also the range where the table should be placed otherwise the previously-generated result table will be overwritten. In general it is recommended to use a new worksheet each time for the newly generated result tables. This would be more suitable because it will provide a better overview and better handling of the results. The only time the second option (existing worksheet) is preferable is when there is a need to have two tables side by side for comparison reasons. In order to ease the handling of the results, a name should be given to each of the created worksheets to allow identification of the type of information stored in it.

## Window 12



Now select 'layout'.

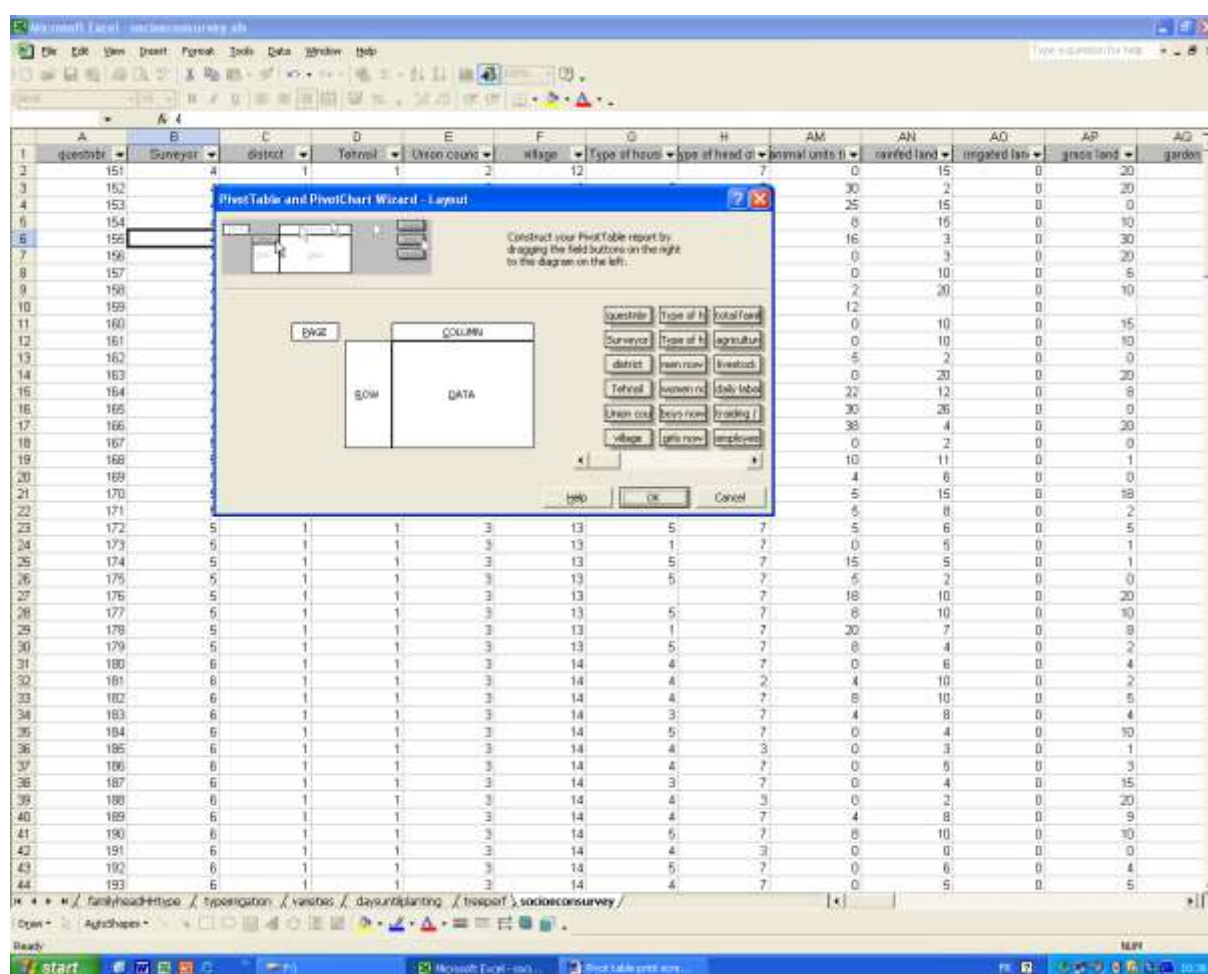
- In the new window, as shown in Window 13, the frame of a table appears with small boxes containing the headings of the columns of the database.

This is the most challenging part of the data analysis because it is necessary to remember what type of data was entered under each header and the type of information required from them. It is possible to obtain the information directly for averages, sums, maximum, and minimum. It is possible to obtain the information indirectly for percentages by comparing the frequencies with the overall number of entries. The frequency can be obtained by counting numbers for specific entries or calculating the sum from Yes/No questions. Here it is necessary to remember which type of calculations the various types of questions allow.

On the right side are all the parameters represented (heading of the columns of your datasheet), and on the left side there is the table that is being created.



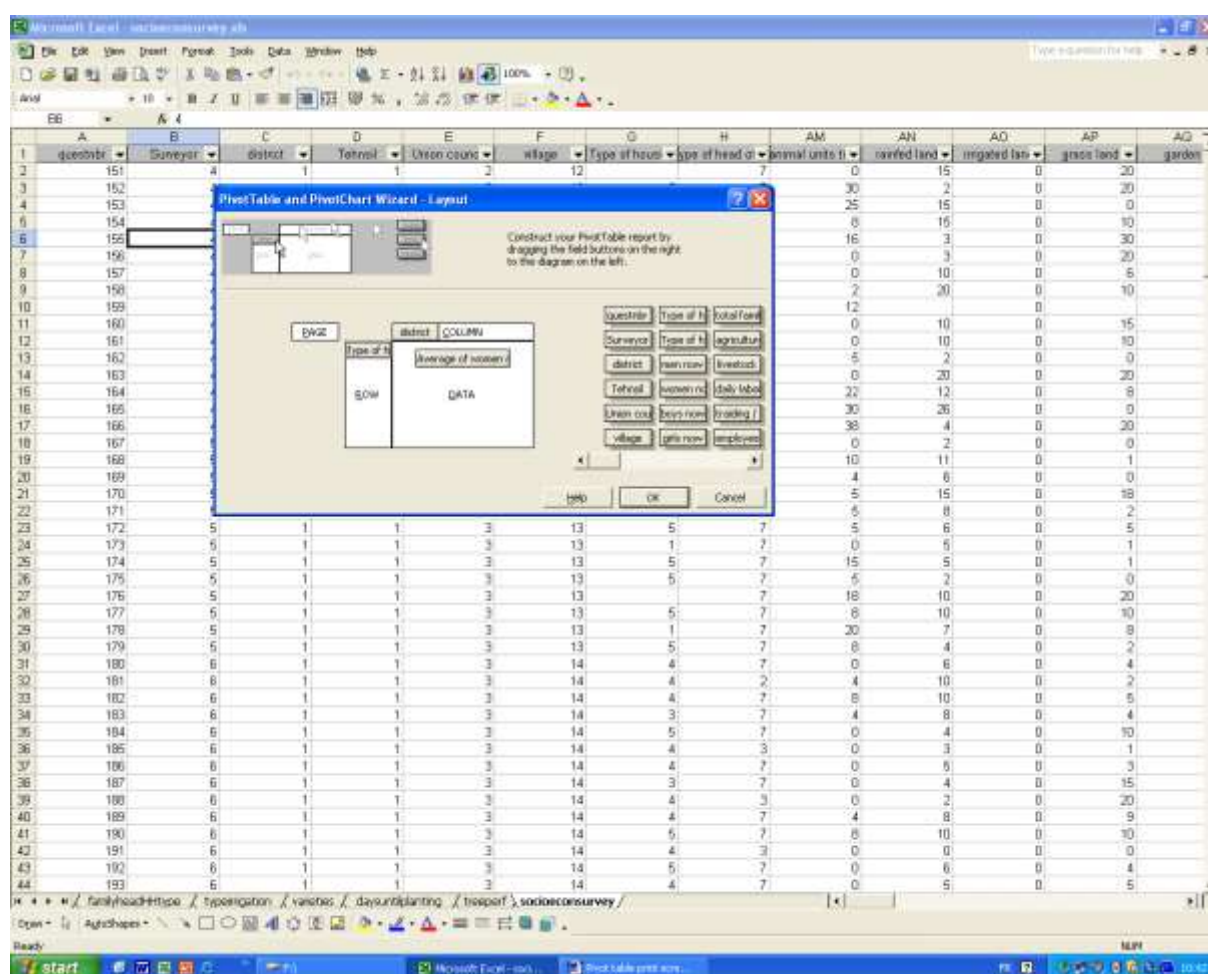
## Window 13



- There are three spaces to where the parameter boxes can be transferred: 1. heading of a column, 2. heading of a row, or 3. as data to be analysed in the central part of the table frame. In the space of columns or rows, the headings with two types of datasets can be transferred: 1. the differentiating parameters (one choice or selection out of a given list) which will split the analysed data in the centre into different categories of the selected parameters, or 2. Ranking parameters (where four categories are possible: poor, fair, good, excellent) in order to calculate the frequency of those categories. It is recommended to transfer only one of these parameters each time, otherwise the tables will become complicated for data interpretation. Therefore, for each of the differentiating parameters it is best to create a new table. Looking into the data within the various columns, there should be a significant difference compared to the overall results (last column or row in the table), to ensure that the selected differentiating parameter has an influence on the analysed data. Aside from transferring parameters into columns or rows, it is also necessary to transfer data into the central part as shown in Window 14. The parameters to be transferred will be mainly quantitative data, but also yes/no responses, or when the number of entries need to be counted, such data can be transferred as well.

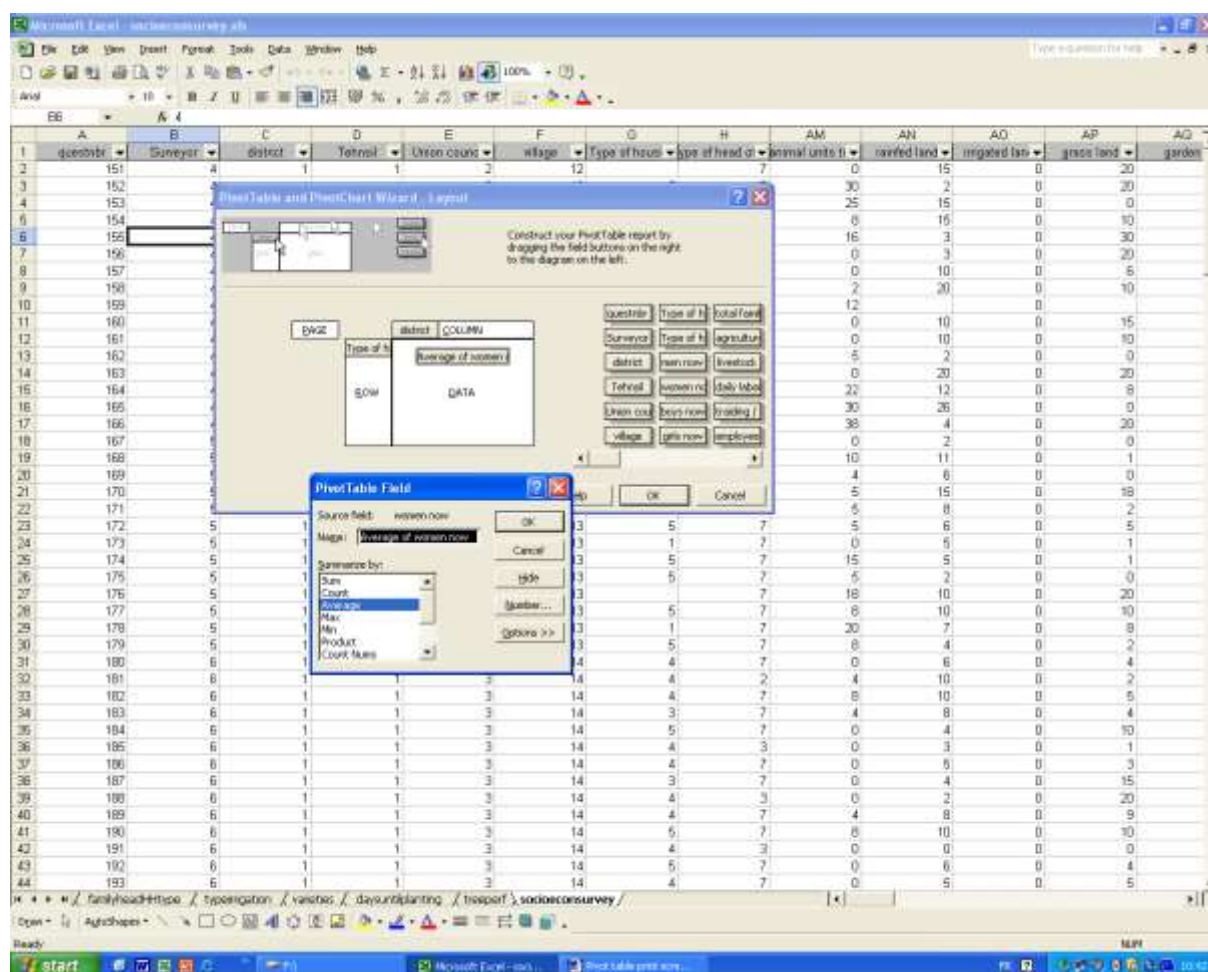


## Window 14



- Once a parameter has been transferred into the data section by double clicking on the same box, a small window pops up, as shown in Window 15, where the different possibilities of functions that data can be subject to or what statistical analysis could be done with the entries can be seen. Select one of the following options: sum, count (counts all types of entries and not only numeric once), average, maximum, minimum, product, count numbers (counts only numeric entries), and lastly four statistical options (two types of standard deviation or variance). These last four options are not really useful for the type of results needed for reporting.

## Window 15



Depending on the result desired, the possibility exists to choose one of the above functions shown in Window 8. In case more than one function is needed it is necessary to transfer again the same parameter into the central part of the table and double click again on the box and select this time the other needed function.

Remember the options for the data analysis depending on the type of question explained in the previous questionnaire development phase.

For a Yes/No question, where Yes was entered as 1 and No as 0, the sum will give the number of the total entries of Yes responses. By comparing it with the total number of entries (count numbers), a percentage can be calculated.

The numeric entries resulting out of questions requesting quantitative information can be subject to the following functions: Max, Min and/or Average.

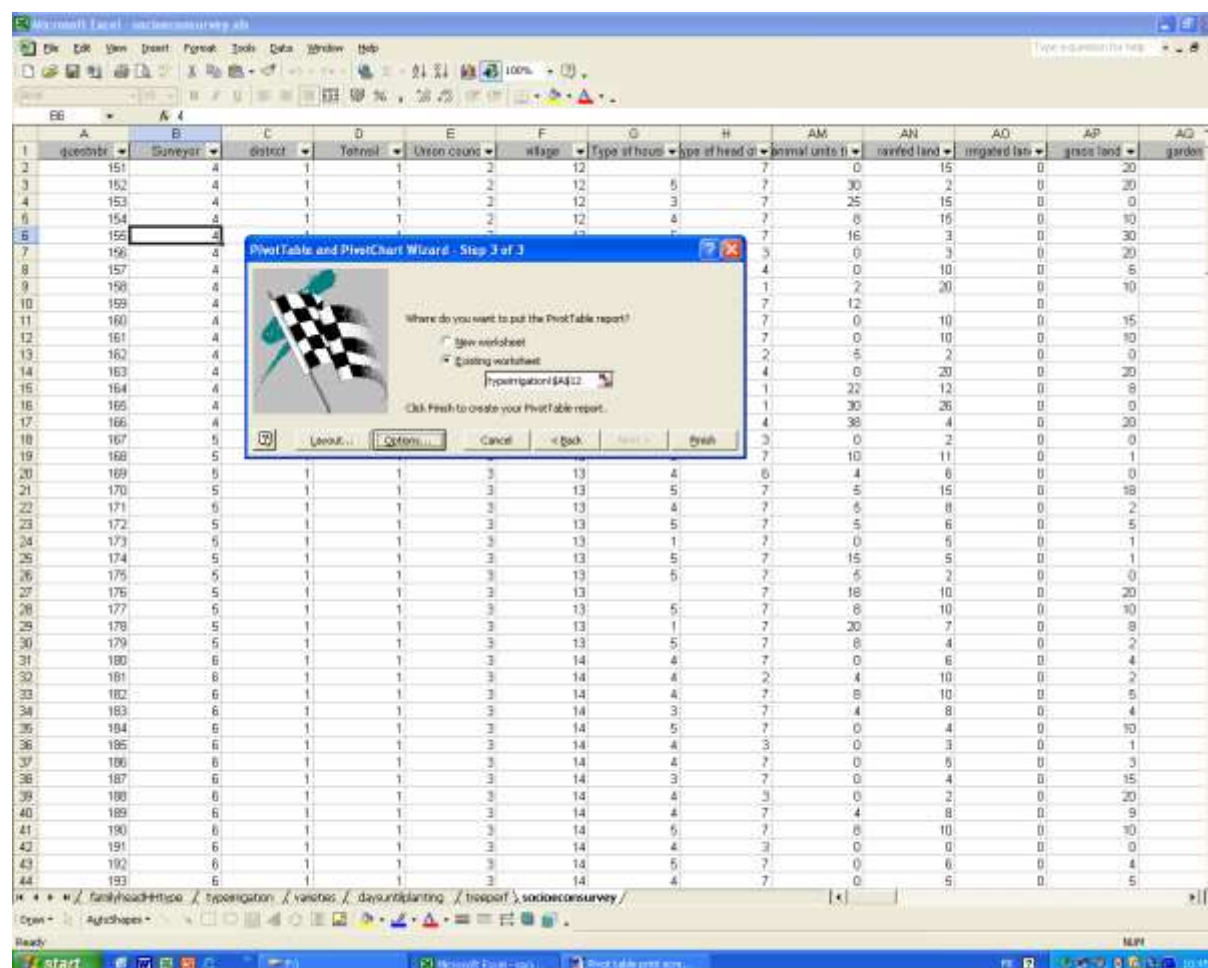
In addition, depending on the type of average which the user is expecting to calculate, some adjustments have to be made beforehand. Either calculates the overall average of a specific entry like debts (this takes into consideration also those households without debts as '0' has been entered into the dataset for those households without debts). If interested only in the average amount of debts from the households with debts all the '0' should be removed in the dataset under the debt amount column.

Finalize the part of creating the whole table with clicking on 'OK'.

It is always better to make several tables. This is easier later on for interpretation and would be more suitable for the establishment of graphs than a single table with too much information.

- Thereafter the same window reappears, shown in Window 16, from which the layout was selected. Check here as to where the results should be placed, either in a new or in an existing worksheet, if not already done. The best option is to choose each time a new worksheet to eliminate risk of overlapping or erasing previously created tables. Now click on the box 'finish'.

**Window 16**



- See in the following window the table of results shown in a simple Pivot Table, see Window 17.

## Window 17

The screenshot shows a Microsoft Excel window titled 'socialsurvey.xls'. The pivot table is set to show 'Average of women row' by 'district'. The data table below it shows the following values:

Type of head of HH	1	2	3	4	5	6	7	Grand Total
1	1.333333333	1.333333333						1.333333333
2		1.75						1.75
3		1.375						1.375
4		1.3						1.3
5		1						1
6	1.333333333	1.333333333						1.333333333
7	1.619047619	1.619047619						1.619047619
Grand Total	1.538888889	1.538888889						1.538888889

Based on these various tables created, it is now possible to make particular tables and graphs required for reporting purposes.